Interaction between Reference Frames, A Concern in Embedded Training?

PMB Sandor[1,2] MD-PhD, D Hartnagel[1] PhD, L Bringoux[2] PhD, C Bourdin[2] PhD, M Godfroy[3,4] PhD, C Roumes[1] MD-PhD

[1] IRBA, BP 73, 91223 Brétigny sur Orge (France), [2] ISM, UMR 6233 CNRS & Université de la Méditerranée, Marseille (France), [3] San Jose State University Foundation, San Jose, CA, USA, [4] NASA Ames Research Center & Human Systems Integration Division, Moffett Field, CA, USA

IRBA, *BP 73, 91223 Brétigny sur Orge (France)*
psandor@imassa.fr, phone+33169237866

ABSTRACT

One of the embedded training goals is to propose a more demanding task generated by a simulation system during a real life routine activity. The systems using virtual reality (VR) can manipulate structural, semantic and localization information and recreate a new environment replicating and/or enhancing the real world context. Then simulation awareness might interfere with perception of reality, especially when body attitude and/or external environment position lead to conflictual inputs. Multisensory space perception studies strongly support the hypothesis of different reference frames for perception and for action as well. We report here experiments examining (i) the effect of structure of the visual field (Bringoux et al., 2009, Perception 38(7) 1053–71), (ii) and dissociation between visual and auditory references (Hartnagel et al., 2007, Perception 36(10) 1487–96) in two research paradigms manipulating different reference frames. Exp 1: A tilted visual frame is known to influence the perception and the restitution of the visual vertical (`rod-and-frame effect' or RFE). The rod-and-frame test (RFT) has been reproduced in an immersive environment (CAVE-like). The RFE observed in this VR environment was qualitatively similar to that obtained with a real visual display but was significantly influenced by the structure of the visual scene and by the adjustment condition (visual control >> visuo-kinaesthetic control). Exp 2: Audio – visual (AV) fusion (i.e. the perception of unity resulting from visual and auditory stimuli under condition of spatial disparity) depends both on the direction and eccentricity of the bimodal stimulus (Godfroy et al., 2003, Perception 32 1233–45) and is assumed to be determined by visual and auditory references (Roumes et al., 2004, Perception 33 Supplement, 142). The contribution of ego- and allocentric-visual cues was further investigated using a fusion task according to 2 gaze positions in total darkness. Results showed that AV fusion in darkness cannot entirely be built on the visual reference and confirmed the hypothesis that the reference frame of the bimodal space is neither head-centred nor eye-centred, but represents rather a compromise. These results are discussed in terms of dynamic combination between coexisting reference frames for spatial orientation and action as it might happen in fully embedded and appended systems.

Report Documentation Page	*Form Approved* *OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **OCT 2009**	2. REPORT TYPE **N/A**		3. DATES COVERED **-**
4. TITLE AND SUBTITLE **Interaction between Reference Frames, A Concern in Embedded Training?**			5a. CONTRACT NUMBER
			5b. GRANT NUMBER
			5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)			5d. PROJECT NUMBER
			5e. TASK NUMBER
			5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **IRBA, BP 73, 91223 Brétigny sur Orge (France)**			8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSOR/MONITOR'S ACRONYM(S)
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release, distribution unlimited**			
13. SUPPLEMENTARY NOTES **See also ADA562526. RTO-MP-HFM-169 Human Dimensions in Embedded Virtual Simulation (Les dimensions humaines dans la simulation virtuelle integree)., The original document contains color images.**			

14. ABSTRACT

One of the embedded training goals is to propose a more demanding task generated by a simulation system during a real life routine activity. The systems using virtual reality (VR) can manipulate structural, semantic and localization information and recreate a new environment replicating and/or enhancing the real world context. Then simulation awareness might interfere with perception of reality, especially when body attitude and/or external environment position lead to conflictual inputs. Multisensory space perception studies strongly support the hypothesis of different reference frames for perception and for action as well. We report here experiments examining (i) the effect of structure of the visual field (Bringoux et al., 2009, Perception 38(7) 105371), (ii) and dissociation between visual and auditory references (Hartnagel et al., 2007, Perception 36(10) 148796) in two research paradigms manipulating different reference frames. Exp 1: A tilted visual frame is known to influence the perception and the restitution of the visual vertical ('rod-and-frame effect' or RFE). The rod-and-frame test (RFT) has been reproduced in an immersive environment (CAVE-like). The RFE observed in this VR environment was qualitatively similar to that obtained with a real visual display but was significantly influenced by the structure of the visual scene and by the adjustment condition (visual control >> visuo-kinaesthetic control). Exp 2: Audio visual (AV) fusion (i.e. the perception of unity resulting from visual and auditory stimuli under condition of spatial disparity) depends both on the direction and eccentricity of the bimodal stimulus (Godfroy et al., 2003, Perception 32 123345) and is assumed to be determined by visual and auditory references (Roumes et al., 2004, Perception 33 Supplement, 142). The contribution of ego- and allocentric-visual cues was further investigated using a fusion task according to 2 gaze positions in total darkness. Results showed that AV fusion in darkness cannot entirely be built on the visual reference and confirmed the hypothesis that the reference frame of the bimodal space is neither head-centred nor eye-centred, but represents rather a compromise. These results are discussed in terms of dynamic combination between coexisting reference frames for spatial orientation and action as it might happen in fully embedded and appended systems.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	**SAR**	**18**	
unclassified	**unclassified**	**unclassified**			

1.0   INTRODUCTION

Embedded training simulation appears to be the way to maintain and spin out crew's operational capabilities when being abroad and far from training facilities, while facing everyday real life in real mission. These missions being mostly routine situation in a quiet non-hostile environment, they may offer the possibility to inject alternative virtual situation during these usual and scheduled phases. Nevertheless introducing such an additional task during a real mission, especially when in a terrestrial or aerial vehicle, necessitates to closely collect differences between this situation and usual training conditions, and check the influence on actual safety and consequences on training results.

The other issue is also to consider that training is developed and performed in laboratory harmless conditions, and that lab condition still does not perfectly replicate the sensory experience provided by the real world, hence the interest of embedded virtual simulation.

The systems using virtual reality (VR) can manipulate structural, semantic and localization information and recreate a new environment replicating and/or enhancing the real world context. More widely, "finality of VR is to give possibility to one or several persons of a sensory-motor and cognitive activity in an artificial world, digitally created, may be imaginary, symbolic or a simulation of some aspects of the real world (RW)" (Fuchs *et al.*, 2006).

So it appears that despite the mixture of VR and RW, sensori-motor and cognitive activity must be efficient considering the training aims, and safe considering actuality.

Then simulation awareness might interfere with perception of reality, especially when body attitude and/or external environment position lead to conflictual inputs. Multisensory space perception studies strongly support the hypothesis of different reference frames for perception and for action as well. Being in a moving vehicle and performing an embedded training session will induce fluctuation of point of interest in space, hence will question this hypothesis. We report here 2 experiments examining (i) the effect of structure of the visual field (Bringoux *et al.*, 2009) in a virtually tilted environment, (ii) and dissociation between visual and auditory references (Hartnagel *et al.*, 2007,) in two research paradigms manipulating different reference frames.

2.0   EXP 1: REFERENCE FRAMES AND SUBJECTIVE VERTICAL ESTIMATE

A tilted visual frame is known to influence the ability to perceive the subjective vertical (SV) without any reference of verticality. A rod that was to be aligned with the gravitational vertical is actually displaced towards the tilted visual environment. Tilted scenes containing objects and also simple frames generate this misperception of the SV. The classical experimental setup devoted to point out and quantify this point is the 'rod and frame test' (RFT), which requires SV judgments in front of a visual square frame tilted at different extents (Oltman, 1968). Today, VR displays may be used to present 3D scenes at remote distance. By tilting of these scenes it becomes possible to study combined influence of scene structure and mode of SV adjustment.

Head-mounted displays, although enabling to create 3-D visual information, suffer most of the time from a reduced field of vision and from the residual presence of a head-fixed visual frame which might concurrently influence the perception of verticality (Mars *et al.,* 2004). Projection-based immersive virtual environments with larger fields of vision have been recently developed and used to manipulate and study spatial orientation.

The first aim of the present study was to investigate whether the tilt of a large-scale immersive virtual environment could yield similar effects on the judgment of verticality as the tilt of real visual surroundings. In other words, we addressed the question of comparability of the indicated SV between real

and virtual worlds. This first step would allow us to validate the immersive virtual environment as a powerful tool for studying the perceived orientation of objects in structured visual surroundings.

The second aim of the present study was to manipulate the 3-D structure of visual surroundings during SV estimates in order to characterise the implication of geometric and polarised features in vertical shifting.

The third aim was to question the role of the adjustment mode. The classical SV setting, by mean of a rod controlled by a remote system was compared to a visuo-kinesthetic (VK) condition, where the participant was holding a rod in his/her hand and could see it during the task.

2.1   Method

2.1.1 Apparatus

Two distinct setups were used to test RFT on SV.
-The first one is a replication (figure 1) of the RFT portable apparatus developed by Oltman (1968). It is composed of a box (57 cm deep X 31 cm wide X 31 cm high) made of wooden white surfaces whose inside edges and corners were marked by black painted lines. The interior of the box was illuminated and the entire device could be tilted by the experimenter at different roll orientations.
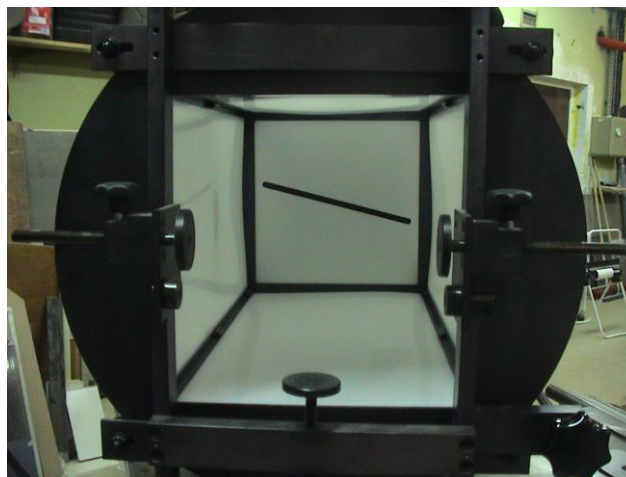

Figure 1: the classical Rod and Frame test (RFT).

A black rod (30 cm long; apparent size: 29.5 deg), fixed to the centre of a black square frame (apparent side size: 30.5 deg), could be independently rotated by the subjects and the experimenter via distinct hand-levers. A protractor, displayed on a disc mounted at the rear of the box and visible only to the experimenter, indicated the deviation of both the frame and the rod from vertical (measurement accuracy: 0.2 deg). Each subject was seated so that his/her face was aligned with the front edge of the box (not seeing the outer environment), and the eye level coincided with the axis of rotation. Subjects were required to keep their heads upright during the adjustment.

-The second setup manipulated in the present experiment is the immersive virtual-reality display (CAVE-like) housed in the Mediterranean Virtual-Reality Centre at Marseilles (France). It is constituted of a 3 m deep X 3 m wide X 4 m high cubic space, with three vertical screens for walls and a horizontal screen for the floor. The three vertical surfaces were back-projected and the ground received direct projection. Stereoscopic separation between left-eye and right-eye images was ensured by colorimetric separation. Filters were installed in the projectors, and subjects were wearing glasses with the same filters for high-quality passive stereopsis. An anti-aliasing mode of projection was used in order to avoid any directional

cue mediated by pixel alignment. An ArtTrack© head-tracking system, featuring infrared recognition of passive markers placed on the glasses, was used to record the subject's head position and orientation (accuracy: 0.05°), and to update in real-time the stereoscopic images in relation to the subject's point of view. Subjects were seated in the immersive environment 2 m away from the front wall. Their field of vision was thus entirely stimulated by the visual display (the apparent size of the virtually projected rear frame reached 73 deg). Subjects were randomly presented with three different virtual scenes (figure 2).

Scene 1   Scene 2   Scene 3
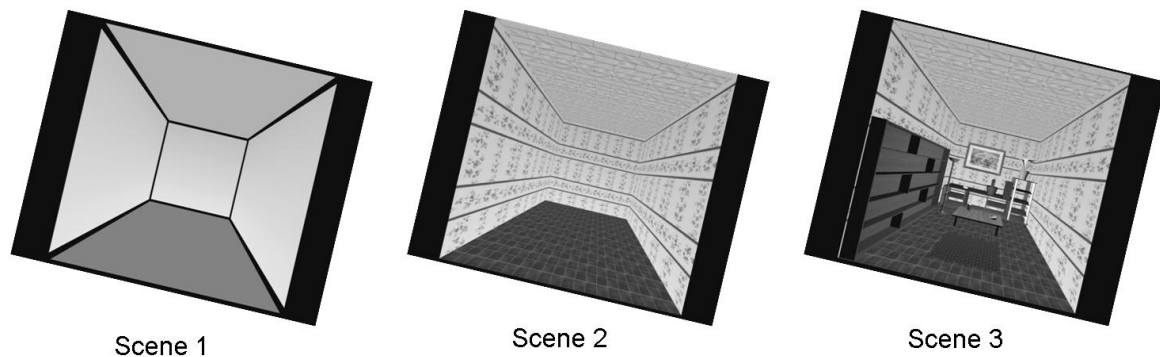
Figure 2: The 3 types of scenes presented through the CAVE system

Scene 1 typically reproduces the RFT environment (with a much larger scale, however). Observers faced a 3 m X 3 m traditional square frame, being immersed in a tiltable cubic space bounded by contrasted orthogonal lines.

Scene 2 consists of an empty coloured wall-papered room with structured floor and ceiling. Features of the scene essentially reinforces the geometrical cues with increased parallel and orthogonal visual lines.

Scene 3 corresponds to a fully furnished room. Virtual furniture included a red bookshelf, a desk with books and green plants, a halogen lamp, a well-known painting by Cezanne attached to the front wall, and a coffee table with a can of soft drink and an ashtray. These elements, lying at different distances from the subjects, added depth cues to the display and were also assumed to enhance high-level (ie cognitive) polarity cues for up and down (Howard and Childerson, 1994).

SV judgments were assessed in two ways (figure 3). In the first (V mode) SV adjustment mode, subjects were asked to set a virtual rod to vertical by means of a computer mouse controlling its orientation in roll. The very small amplitude of mouse displacements (<1.5 cm) could not yield accurate information about the angular motion of the rod. The projected rod was centred relative to subjects' eye level, at a distance where it could be held with the extended arm.

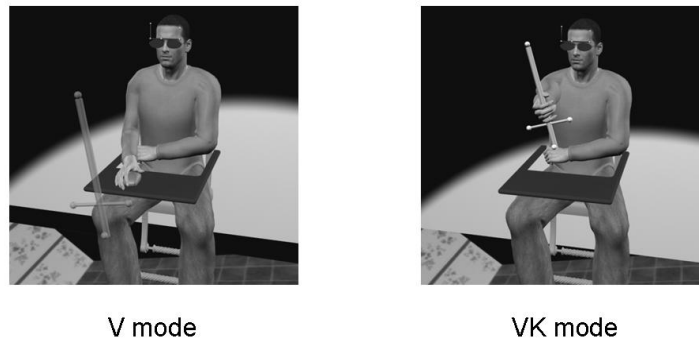V mode   VK mode

Figure 3: The 2 modes of SV judgement in the CAVE system, visual (V) and visuo-kinaesthetic control (VK). Subject is wearing stereo glasses.

Virtual rod features (colour, apparent size, distance to the observer, projected height) were computed on the basis of the characteristics of a real hand-held rod (used for the second adjustment mode visuo-kinaesthetic VK) measured for each subject before starting the experiment. In V mode, the rod orientation was controlled only by visual inputs. In VK mode the subject held a light plastic rod (40 cm long; 1 cm in diameter; weighing 60 g with uniform mass distribution) in the right dominant hand, and to adjust it along the vertical axis with an extended arm. Subjects were instructed to keep the centre of the rod at eye level and to look at it during adjustment. Markers positioned on the rod enabled us to continuously record its orientation via the ArtTrack© system (measurement accuracy: 0.05°), and ensured that the final location of the centre of the rod was kept around the same position across the trials. In VK mode, both visual and kinaesthetic inputs allowed the subjects to control the rod orientation.

2.1.2 Procedure

The experiment was divided into two counterbalanced sessions, corresponding to the two adjustment modes manipulated in the immersive environment. Before the first session, subjects were required to perform SV judgments through the portable RFT. Specifically, they were asked to "align the rod along the gravity axis" by rotating the hand lever. Nine frame tilts (-38°; -28°; -18°; -8°; 0°;+8°;+18°; +28°; +38°, positive to the right) and four initial rod orientations (-45°; -25°; +25°; +45°) were manipulated to define basic individual profiles. Pseudo-random presentations of initial rod positions and frame tilts were counterbalanced in order to cancel any order effect.

2.1.3 Data processing

Final SV adjustments were collected during the RFT and averaged for obtaining mean individual signed deviations relative to the gravitational vertical (constant errors) for each scene tilt. Rod and head location as well as orientation in 3-D were monitored by the tracking system throughout each trial in the immersive environment. Mean signed and unsigned deviations of the rod relative to the gravitational vertical (constant and variable errors, respectively) were processed for characterising SV judgments in the immersive environment. Differences between "raw" SV adjustments were tested by a repeated measures multifactorial analysis of variance (ANOVA) conducted on the mean signed deviations of the rod relative to vertical.

The ANOVA factors were scene tilt (9 levels), visual scene (3 levels) and adjustment mode (2 levels).

2.2   Results

2.2.1. SV in real vs. virtual environment

As illustrated in figure 4, the SV appeared as a sinusoidal function of the scene tilt from -38° to +38° in both the RFT and the immersive environment. This shape is typical of classical RFT SV shift effect reported in the literature. The correlation between the mean data recorded in the RFT and the immersive environment was high and significant.
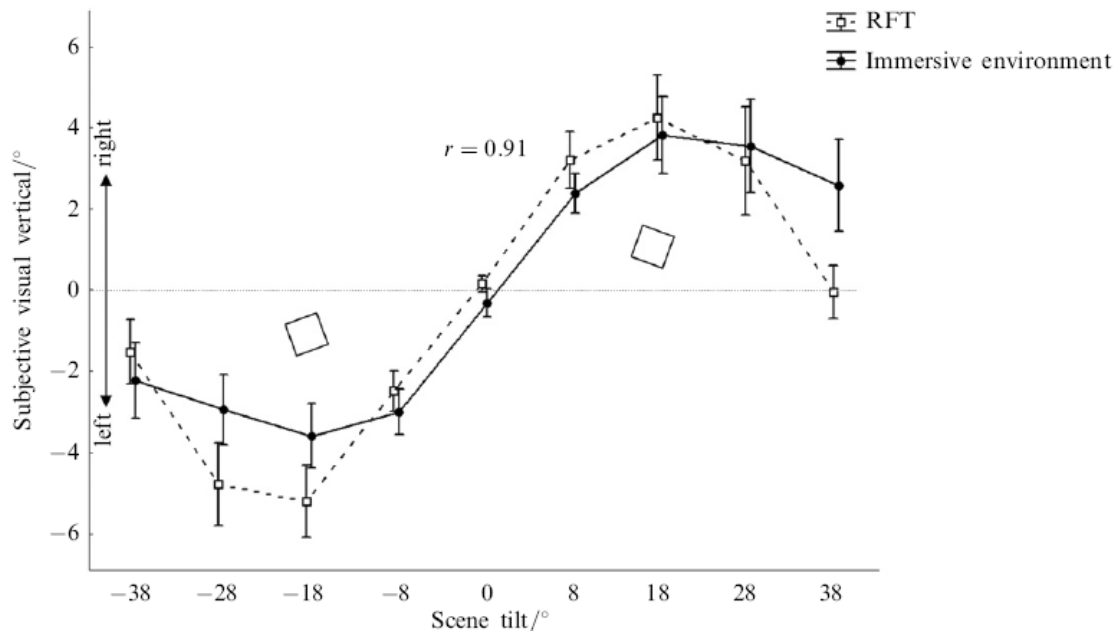

Figure 4: Mean SV settings as a function of tilt of the visual scene in real (portable rod-and-frame test, RFT) and the virtual immersive environment. Sinusoidal curve is typical of a classical rod-and- frame effect (RFE). Error bars represent 95% confidence intervals.

2.2.2. SV shift in immersive environment.

In line with the previous results, the ANOVA conducted on the mean signed deviations of the rod relative to vertical revealed a main effect of scene tilt ($F_{8,224} = 59,40$, $p < 0,001$). This confirms the SV shift effect in the immersive environment, whatever the experimental condition.

2.2.2.1. Influence on SV shift of the visual scenes in immersive environment.

A significant interaction was found between scene tilt and visual scene ($F_{(16,864)} = 4.05$; $p < .001$). It shows that the RFE increased as a function of the structure of the visual scene (figure 5). The RFE magnitude was larger in scene 3 than in scene 2 ($p < .001$), and was even larger in scene 2 than in scene 1 ($p < .001$).
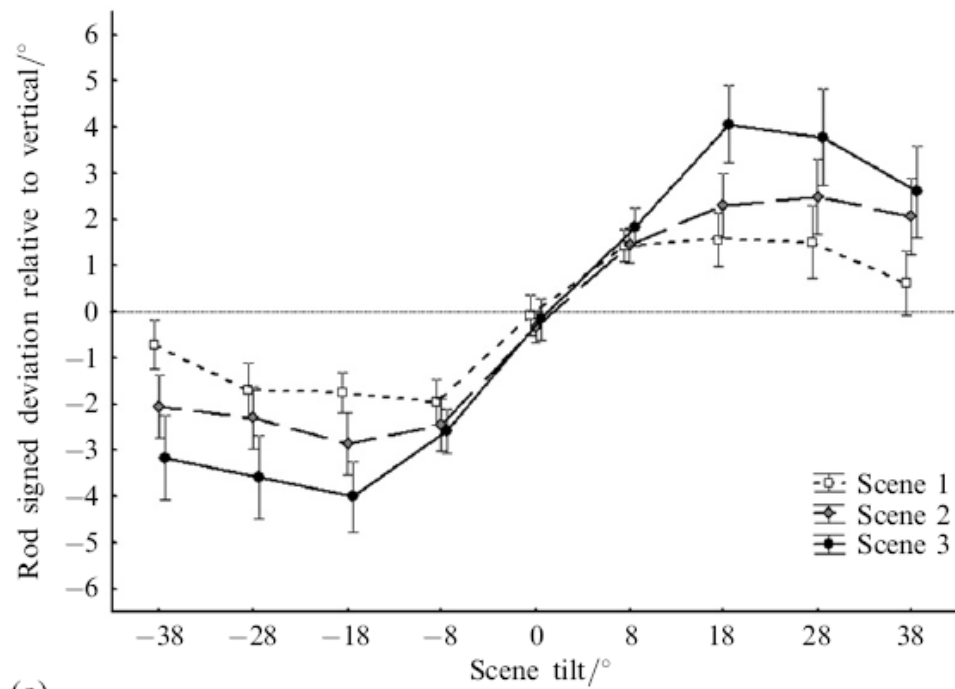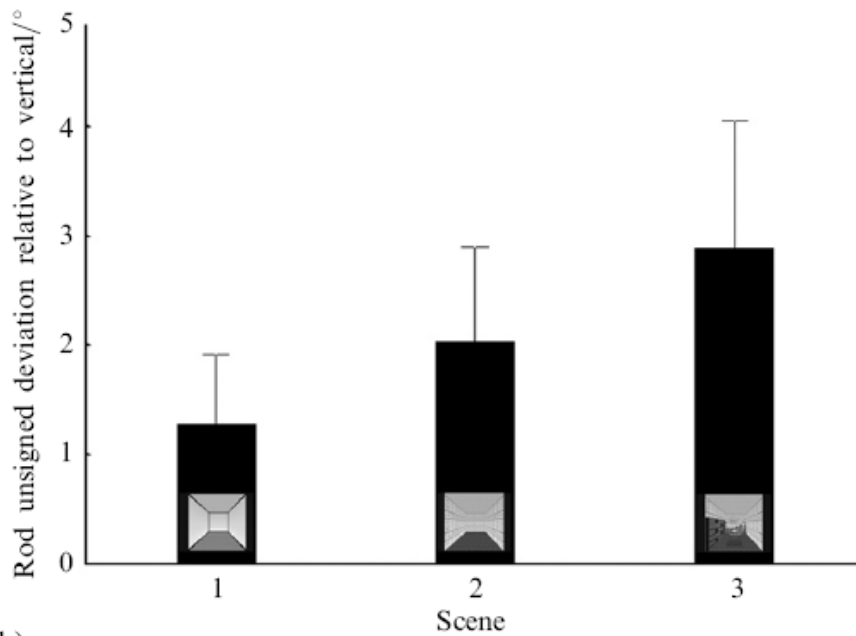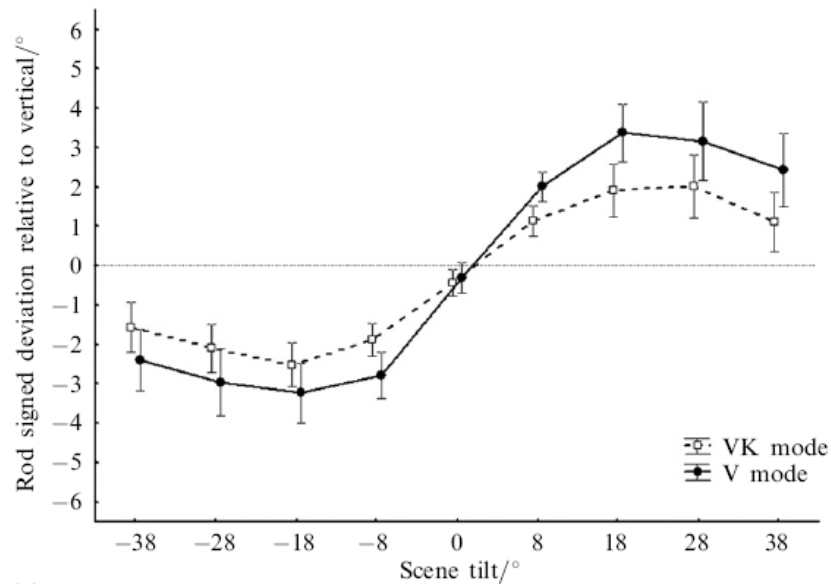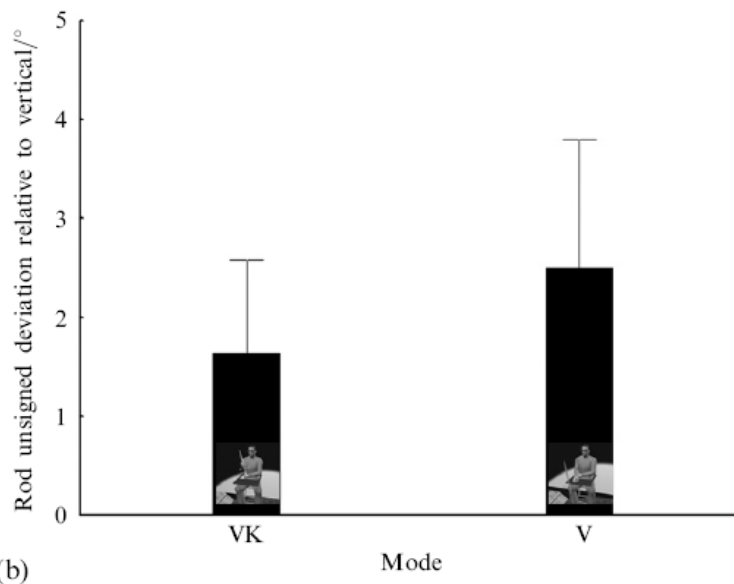
Figure 5: Mean signed (a) and unsigned (b) SV deviation as a function of tilt and structure of the visual scene. The more structured the scene, the greater the rod-and-frame effect.

2.2.2.2. Influence on SV shift of the adjustment mode.

A significant interaction was also found between scene tilt and adjustment mode when comparing the mean signed deviations of the rod relative to vertical ($F_{8,224}=12{,}75$, $p< 0.001$). SV deviation was greater for adjustments performed in V mode than in VK mode (figure 6).

(a)

(b)

Figure 6: Mean signed (a) and unsigned (b) SV deviation as a function of tilt and adjustment mode. Deviation appears greater when assessed under visual control (V mode), to visuo-kinaesthetic control (VK mode). This effect is even more apparent on mean unsigned deviations of the rod relative to vertical ($F_{1,28}=20.48$ $p < 0{:}001$

2.2   Discussion

Tilting of a scene is known to induce a shift towards the tilt side of the subjective judgment of the vertical. It can be shown and quantified by the RFT test, which is a real almost wood-made small sized device used since 40 to 60 years. This effect is maximal between 18° and 28 °tilt. This experiment validates the possibility to transfer this effect from a small sized RFT to a wide immersive virtual CAVE, as SV deviation appears similar when CAVE scene is poorly structured. The second point is the fact that SV is highly influenced by enrichment of the surroundings; the more the scene is structured, the more SV may be deviated. The third point is that SV deviation in VK mode is smaller than in V mode.

As an applied consequence of these observations in the purpose of this meeting, we may argue that the visual surrounds may influence vertical perception, which is one of the main features used to pilot vehicles, especially airplanes. Of course, the classical concern about misperception of vertical (and hence horizontal) is sensory illusion, followed by erroneous flight control, which can be avoided by autopilot using and/or following strict instrument flying rules. Sensory illusion may be generated by a short time movement (e.g. head movement during a turn) or constructed along several minutes (e.g. leans effect, Gillingham and Wolfe, 1986) that are recovered through training. Introducing an additional cognitive task including presentation of strongly structured visual information may induce an additional level of reference shift. As in our results, it appears that VK mode is less influenced than V mode, it should be of interest to preserve during an embedded training session an actual motor tasks of the crew, in order to keep an active reorientation responsibility. For safety, this is to be separated from the flying task, which should remain devoted to an autopilot or to others crewmembers.

SV determination is proposed to emerge from contribution of different reference frames (RF), defined as a system of coordinates, set of axes used to code and update position and orientation in space. Usually referred are the geocentric RF (direction of gravity, vestibular input), egocentric RF (body axis, proprioceptive input) and allocentric RF (surroundings, visual input). None of those is fully responsible for the present results and a combination may be inferred. The resulting RF should be the weighted sum of the different RF. This weighting should be dynamically performed at high central nervous system (CNS) level, and includes a multisensory integration as described in posture control (Oie *et al.*, 2002), which also refers to vertical information. These authors expressed the weighting process being intra-modal (surroundings motion amplitude on vision, touch motion amplitude on touch) and inter-modal (touch on vision). Although our experiment differs as subjects were immobile in a stationary environment, description of "in-line" control of RF weighting process with the reinforcement of the allocentric weight when scene structure is reinforced should be similar. A relation to neurophysiological supports related to different RFs (Commitieri *et al.*, 2004) give support to the understanding how final result is obtained, and explained how geo and ego-centric RF, that need to be updated frequently, may be exceeded by allocentric RF, that describes relationship between objects in space. This should be a way to shift from piloting to navigation. Introducing an additional task in embedded training must be done considering preservation of this feature, i.e. preservation of the allocentric RF once constituted.

Dynamically updated information from the surroundings is obvious during piloting, and even more if an additional cognitive task is added. To facilitate information capture or increase information rate one option is to propose multimodal coherent information in order to reinforce the intelligibility and simplicity of presentation (Cornell *et al.*, 2002). Co-presentation of spatialized auditory and visual stimulation is likely. This presentation may constitute an additional layer of RF to those already used and weighted for self orientation in space. The following second experiment will address theses RF.

3.0 EXP 2: AUDIO-VISUAL FUSION AND GAZE DIRECTION

Perception of the world is basically multisensory, consisting of a unified perception of the various unimodal inputs each being fitted with a particular frame of reference for position in space. When combining visual and auditory cues in space, the brain can tolerate a certain amount of spatial disparity between unimodal parts of the stimulus but recognizes both stimulations as coming from same location in space. This phenomenon is called `perceptual fusion' and is usually investigated through the ventriloquism effect (Jack and Thurlow 1973), in which the perception of the spatial location of a sound is biased in the direction of a visual stimulus (Bertelson and Radeau 1981).

Audio-visual (AV) fusion has been investigated by Godfroy *et al.* (2003) to evaluate the perception of unity of spatially disparate stimuli in a frontal 2 dimensional frame. They showed that AV fusion capability varied with the horizontal eccentricity of the bimodal stimulus. They determined fusion areas for a set of loudspeakers spread over the frontal visual field by estimating the spatial extent over which a luminous spot could be displayed and still be perceived as confused with the sound in over 50% of the trials. The smallest fusion areas were encountered in the median saggittal plane (MSP), and AV fusion areas were found to be symmetrical in relation to that plane (figure 7).
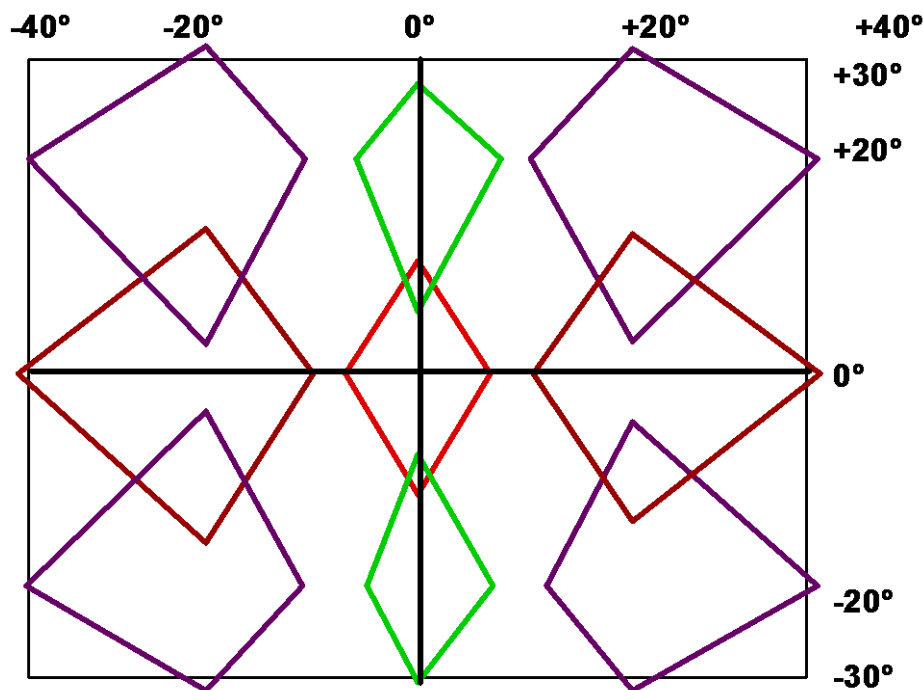

Figure 7: Visual-Auditory fusion areas (from Godfroy *et al.*, 2003):

These variations in fusion areas over space followed closely the spatial resolution of the auditory system (Oldfield and Parker 1984). It was postulated that auditory accuracy plays a major role in AV fusion magnitude.

As vision information is initially coded at a retinal level, and thus depends on the eye position, the visual reference frame is considered eye-centred. For audition, spatial information estimation depends mainly on interaural time differences (ITDs), interaural level differences (ILDs) for azimuth, and on spectral cues for elevation (Blauert 1983). These cues vary with head position; therefore, the auditory reference frame is considered head-centred.

In Godfroy *et al.* (2003) experiment, the participants performed the AV fusion task looking straight-ahead; so both reference frames (eye and head) were aligned. The arising question: does gaze direction affect AV fusion in space? Roumes *et al.* (2004) previously reported effects of gaze direction on audio-visual spatial integration (gaze and head aligned or 20° shifted apart). In their experiment AV fusion areas vary with gaze position (i.e. gaze shift alters fusion areas in azimuth). As an explanation the reference frame of AV fusion is neither head-centred (auditory) nor eye-centred (visual) but appears to be an average between the two reference frames. The experiment reported here is a replication of this experiment, but in total darkness, in order to avoid the possible frame effect linked to the edge of the projection area. The aim was to determine whether AV fusion depends only on the relationship between unimodal egocentric reference frames and does not depend on any peripheral allocentric frame.

3.1 Methods

3.1.1 Apparatus

The subject was positioned at the axis of symmetry of an acoustically transparent, hemicylindrical screen, 120 cm in radius and 145 cm in height. The subject's head was stabilised by a custom bite-bar with the eyes at mid-height of the screen. The head and body were rotated 10° left of the screen axis of symmetry to increase the space of investigation when the fixation spot was presented 20° to the right. No alternative right-orientation shift was tested because no laterality effect had been found in the previous luminous experiment (Roumes *et al* 2004). Gaze orientation was monitored with an ASL 504 (50 Hz) eye-tracker located 45 cm in front of the subject at a level lower than the investigated field of view to prevent any visual masking. To eliminate allocentric cues, the experimental room was in total darkness, and noise level was reduced as much as possible (<39 dBA).

A fixation spot was provided by a laser beams (λ=635 nm (red), <1 mW), displayed either straight-ahead or 20° laterally shifted to the right. The bite-bar controlled the head position (i.e. the auditory (A) RF) and the eye-tracker measured the eye position (i.e. the visual (V) RF). A gaze-fixation angular error smaller than 1.66° during a random time varying inside 300 to 700 ms interval was conditional on stimulus presentation. This test controlled initial spatial relationship of A and V RFs at the bimodal stimulus onset.

3.1.2 Stimulus

The bimodal stimulus consisted of a 49 dBA broadband pink noise presented for 500 ms (with 20 ms onset and offset ramps) in synchrony with a V spot (1°Ø, 3 cd.m^{-2} luminance). Angular vertical and horizontal separation control was the purpose of the experiment.

The A part of the bimodal stimulus was delivered by one of the 19 loudspeakers (LS) (Fostex FE103, 10 cm Ø) located behind the screen, oriented toward the subject's head (figure 8). The V part of the bimodal stimulus was provided by a laser beam (λ=532 nm (green), <3 mW). Mirrors movements and laser on-off were noiseless to the subject.
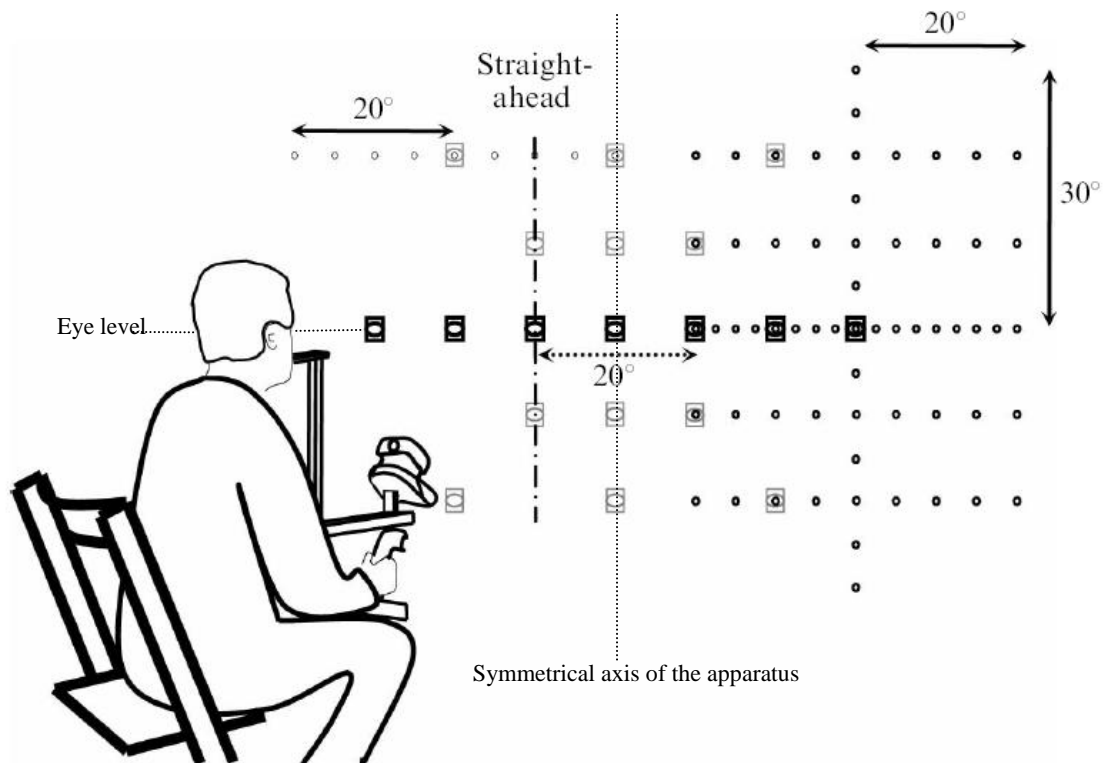
Figure 8: Experimental setup. In total darkness, the subject faced the LS 10° to the left of the symmetrical axis of the apparatus, the head was maintained with a bite-bar, and the gaze was monitored with an eye-tracker. A red fixation point was displayed either straight-ahead or 20° laterally shifted to his/her right. The auditory part of the bimodal stimulus could be produced by one of the 19 LS placed behind the screen in synchrony with a 1° green laser beam. 61 positions of the spot of light were tested for each of the 7 LS at eye level and 9 positions for the remaining LS.

3.1.3. Procedure

The experiment was divided into two counterbalanced modes, corresponding to the two visual fixation modes, corresponding to RF aligned or dissociated, randomly presented in the same experimental session. Subjects were not aware of the following aspects of the procedure. At eye level, AV disparities on ±20° azimuth by 2.5° steps were tested; at ±10° and ±20° elevation levels 5° steps were tested. On each LS located at eye level, 5° steps were tested from -30 to +30° vertical.

All bimodal combinations were repeated 5 times so that each subject responded to 5350 trials, i.e. [LSs at eye level (7) X spot position (61) + LSs above and under eye level (9) X spot position (12)] X reference-frames condition (2) X repetition (5).

3.1.4. Task

The subject had to judge the perception of unity resulting from the bimodal stimulus using a joystick. When the spot and the pink noise were perceived as coming from a unique and common location in space, "fusion" response was selected by pulling the joystick. Otherwise the "no fusion'" response was selected by pushing the joystick. Considering the precautions taken to check eye position, stimulus presentation was gaze-dependent. A trial took about 3 s to complete.

3.1.5 Analysis

A rate of fusion was derived for each disparity tested. Then, fusion limit was estimated with a probit analysis, from the 50% "fusion" response rate. For each of the seven loudspeakers at eye level, 12 limits were estimated, 10 in azimuth (5 on the left and 5 on the right) and 2 in elevation. For the other twelve LSs, only 2 limits in azimuth were calculated. These limits allowed definition of the so-called "fusion area" for each location of a loudspeaker. Statistical analysis was performed on fusion limits for each subject in the two experimental conditions tested (RF aligned or dissociated).

3.2 Results

Only fusion data obtained at eye level will be presented here. Fusion areas were anisotropic as fusion limits in elevation were larger than those in azimuth. The fusion limits in azimuth varied with horizontal eccentricity of the LSs for each condition (figures 9a and 9b). The narrowest fusion areas for the RF-aligned condition were for LSs positioned straight-ahead (6.911°, SD=4.125°) and the fusion limits in azimuth increased with eccentricity of the LSs from the straight-ahead location (figure 9a). In the RF-dissociated condition, the narrowest fusion areas were for LSs positioned between straight-ahead and the fixation point (8.020°, SD. 4.641°) and the fusion limits in azimuth increased with horizontal eccentricity from this intermediate point (figure 9b).
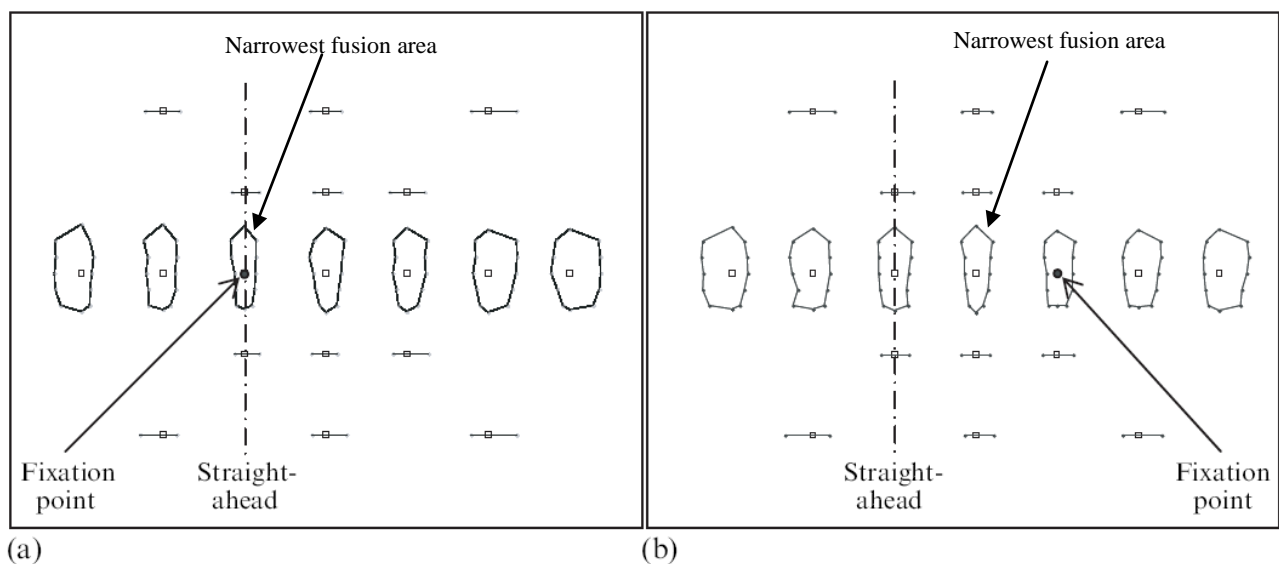

Figure 9: Fusion areas pooled across subjects for all loudspeakers in the aligned-RF condition (a), and in the dissociated-RF condition (b).
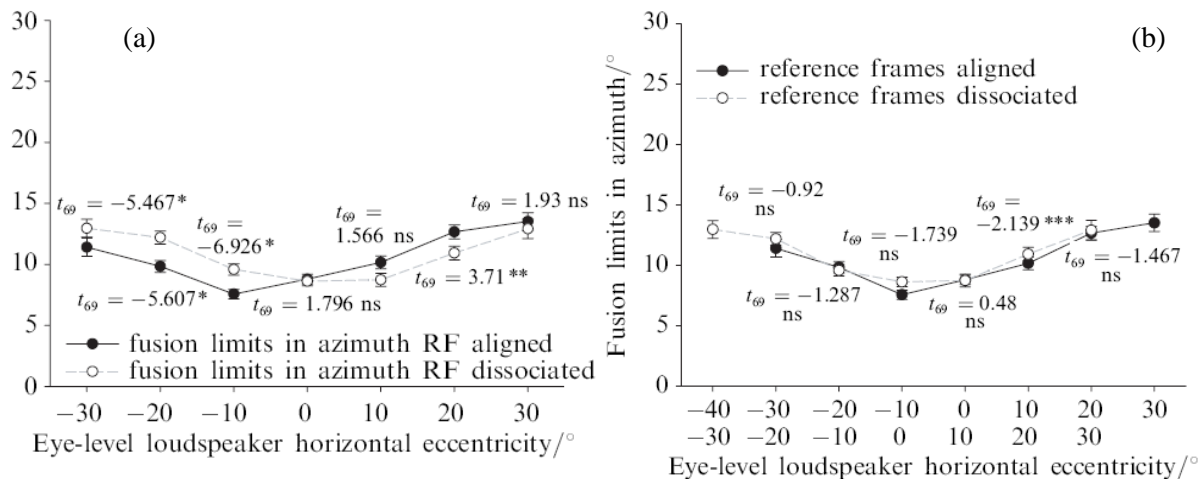
Figure 10: (a) Fusion limits in azimuth when the reference frames are aligned or dissociated for LSs at eye level. (b) Fusion limits in azimuth for LS at eye level, when the dissociated fusion limits values were displaced 10° to the left (*=p<0.001, **=p<0.01, ***=p<0.03).

An ANOVA on fusion limits in azimuth was conducted with horizontal eccentricity of the LSs and RF condition as the 2 within-subjects factors. Statistical analysis showed no significant difference between the two RF conditions ($F_{1,1302}$=1,056, p=0.3). However, AV fusion limits in azimuth varied with the horizontal eccentricity of the LSs ($F_{6,1302}$=19.155, p<0.0001). This variation along the frontal visual field appeared laterally deviated in the direction of gaze shift in the RF-dissociated condition. This effect is confirmed by the significant interaction between RF conditions and horizontal eccentricity ($F_{6,1302}$=4.77, p<0.0001). Differences between the two RF conditions were significant for 4 paired eccentricities out of the 7 tested (figure 10a). When the curves for the RF dissociated condition were 10° left-shifted (e.g. data from LS at -30° of eccentricity in the RF-aligned condition were compared with data from LSs at -20° of eccentricity in the RF-dissociated condition), the matching of the data sets was highly improved and only one comparison was still significant out of six (figure 10b).

3.3   Discussion

The current experiment shows that even though AV fusion closely follows the properties of audition over space, it also depends on the relative position of both unimodal sensory captors. This latter effect is mainly due to egocentric cues as it remains effective in darkness without any visual allocentric biases. In the aligned RF condition, the narrowest fusion areas were in the straight ahead position. Fusion limits in azimuth increased with lateral eccentricity; and fusion areas were symmetrical in relation the straight ahead direction. With these three characteristics, the head-centred RF hypothesis for AV fusion is supported. This result is consistent with a previous experiment in which subjects fixated straight-ahead in congruence with the allocentric visual cues (Godfroy *et al* 2003). However, it differs from previous results from an experiment conducted in light where fusion areas exhibit a slight asymmetry (Roumes *et al* 2004). As seen in exp 1, peripheral visual cues influence judgment. Sheth and Shimojo (2004) argued for the domination of the extrinsic frame of reference in the representation of space. They hypothesised a single undissociated spatial map. Following this assumption, peripheral visual cues provided by the visual display in a luminous environment (Roumes *et al* 2004) may account for the asymmetry of the AV fusion areas relative to the subject straight ahead direction, despite the alignment of the visual and the auditory reference frames. The boundaries of the luminous display may have altered AV fusion over space by providing an allocentric RF.

In total darkness, introducing the dissociated RF condition without allocentric visual information again reveals fusion areas vary with the eccentricity of the LS. But this variation significantly differs from the aligned-RF condition. The narrowest fusion areas in the azimuth dimension are those from the LS located between straight-ahead and the gaze position (i.e. the vertical axis including 0° in azimuth in the apparatus). These latter fusion areas are the most intrinsically symmetrical ones; and all other fusion areas are symmetrically organised on both sides. These results are in line with previous data collected in a luminous asymmetrical-background setup (Roumes *et al* 2004), where allocentric visual cues may have emphasised the symmetrical axis of the visual display. As the same gaze-shift effect was found in the current experiment in darkness, the previous gaze-shift effect can be assumed to be an effect of the egocentric reference frames dissociation. Nevertheless, the overall symmetry of AV fusion areas over space is not entirely determined by gaze position as it only shifts by half the magnitude of the misalignment of the V and A RFs. So, the RF for fusion space can be considered neither eye-centred nor head-centred, but rather as resulting from contributions from both egocentric reference frames. These behavioural data point to a equally weighted dual contribution of unimodal reference frames to AV fusion space in a simple task. The arising question is evolution of this weighting in more demanding tasks.

4. DISCUSSION CONCLUSION

The results of these two experiments made in lab conditions may be taken into consideration in most of embedded training situation whatever the type of vehicles used (airplanes, land vehicles or even vessels). The main problem is about space perception, peculiarly position of body and/or vehicle in space. Perception of body or vehicle in space is well studied in flight because its implication with safety. For this reason we focus discussion on flight situation.

In real flight conditions, systems have been developed to help crew and avoid sensory illusions and decrease basic piloting workload, thus turning the crew into a system manager and most of time a supervisor. In these conditions, pilot have to trust information display more than body sensation. Fully embedded training introduces new deals. To be efficient, Virtual Simulation is supposed to be close to reality, but results of the first experiment showed that the more the details are presented, the more the SV judgment is influenced. This SV judgment ability is on the edge of safety in airplane flying, introducing new incongruent virtual stimulations may produce additional mental load and may be a new source of motion sickness. Moreover possible presentation of fictitious information may induce a blur in the rigid belief into instrument information and create a gap in safety.

Results showed also that the V response condition is more influenced by the tilted frame than the VK response. This result support the hypothesis proposed by Bridgeman (1997) about the dissociation between cognitive and motor spaces. In the virtual embedded environment, operator will have to react in a Visual Virtual world unmatched with reality; in that condition it can be supposed that cognitive space will be asked. Then, subjective vertical will probably be influenced during the embedded training.

Using Audio is supposed to facilitate information processing by reducing workload and improving perception. Vision and audition are the main sources of the allocentric RF for CNS. Experiment 2 suggests that the frame of reference for AV fusion results from a construction, a combination of different sensory information. Taking into account previous results (Roumes *et al.,* 2004) it has been supposed that the structured visual environment is responsible of a slight asymmetry in the results. Even if it would be the case, this influence is very weak, the main point is that multisensory space is not organize on visual space only (cf. gaze position). Dealing with different sensory modalities, the CNS combines different spaces without giving advantage to vision. Taking this assumption it might be possible that conflict between senses may be caused by embedded training.

Altogether, that means that until more experiment will be proceed on the problem of dealing with unmatched realities (Virtual, Augmented Reality) it seems to be worth to advocate that the operator involved into an embedded training session should not be in responsibility of flight safety procedure and decision.

REFERENCES

[1]   Bringoux L, Bourdin C, Lepecq JC, Sandor PMB, Pergandi JM, Mestre D (2009). Interaction between reference frames during subjective vertical estimates in a tilted immersive virtual environment. Perception, 38, pp 1053-1071.

[2]   Bertelson P, Radeau M, (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. Perception & Psychophysics 29 pp 578-584

[3]   Blauert J, (1983) Spatial Hearing. The psychophysics of human sound localization (London: MIT Press)

[4]   Bridgeman B, Peery S, Anand S. (1997). Interaction of cognitive and sensorimotor maps of visual space. Perception & Psychophysics, 59 pp 456-469

[5]   Committeri G, Galati G, Paradis A, Pizzamiglio L, Berthoz A, LeBihan D (2004) Reference frames for spatial cognition: different brain areas are involved in viewer-, object-, and landmark-centered judgments about object location Journal of Cognitive Neuroscience 16 pp1517-1535

[6]   Fuchs P, Berthoz A, Vercher JL (2006). Introduction à la réalité virtuelle. In : Traité de la réalité virtuelle, T1 L'homme et l'environnement (Treaty of the virtual reality, volume 1, Human being and environment, in French). Berthoz & Vercher Eds. Presse de l'Ecole des Mines, Paris, France.

[7]   Godfroy M, Roumes C, Dauchy P(2003), Spatial variations of visual-auditory fusion area. Perception, 32, pp 1233-1245.

[8]   Hartnagel D, Bichot A, Roumes C (2007). Eye position affects audio-visual fusion in darkness. Perception, 36, pp 1487-1496.

[9]   Howard IP, Childerson L, (1994). The contribution of motion, the visual frame, and visual polarity to sensations of body tilt. Perception, 23, pp753-762

[10]   Gillingham KK, Wolfe JW (1986). Spatial orientation in flight. USAFSAM-TR-85-31, pp 80

[11]   Jack CE, Thurlow WR, (1973) Effect of degree of visual association and angle of displacement on the ventriloquism effect. Perceptual and Motor Skills 37 pp 967-979

[12]   Oie KS, Kiemel T, Jeka JJ (2002). Multisensory fusion: simultaneous re-weighting of vision and touch for the control of human posture. Cognitive Brain research, 14 pp164-174

[13]   Oldfield SR, Parker SPA (1984) Acuity of sound localization: a topography of auditory spaces. I. Normal hearing conditions" Perception 13 581-600

[14]   Oltman PK, (1968). A portable rod-and-frame apparatus. Perceptual and Motor Skills 26, pp503-506

[15] Roumes C, Hartnagel D, Godfroy M (2004). Visual –Auditory fusion and frames of reference. Perception, vol 33 supplement ECVP, pp142

[16] Sheth BR, Shimojo S, (2004) Extrinsic cues suppress the encoding of intrinsic cues. Journal of Cognitive Neuroscience 16pp339-350